Volume 1, Issue 1, 2025

The Algorithmic Leader: Cognitive Biases and Trust in AI-Mediated Decision Making within Large-Scale Financial Systems

Y. Xiao[1]

[1]*Strategic Leadership Division, A Major Canadian Financial Institution, Ottawa, Canada (board@ojbs.org)*

Abstract— As artificial intelligence systems permeate large-scale financial institutions, executive decision-making increasingly relies on algorithmic recommendations for risk assessment, portfolio optimization, and regulatory compliance. This paper examines the psychological mechanisms underlying senior leaders' acceptance or resistance to AI-mediated insights within a major Canadian financial organization. Drawing on organizational psychology and behavioral economics, we analyze how cognitive biases—particularly automation bias and algorithm aversion—shape trust dynamics between human judgment and machine-generated predictions. Our investigation reveals that executives oscillate between over-reliance on algorithmic certainty and reflexive distrust of non-transparent models. By integrating theoretical frameworks from prediction machines and digital collaboration research, we propose a balanced governance model that preserves regulatory accountability while leveraging computational advantages in complex financial environments.

Keywords: artificial intelligence; cognitive biases; financial decision-making; automation bias; algorithm aversion; trust calibration; regulatory compliance

1. INTRODUCTION

The contemporary financial services landscape has witnessed an unprecedented integration of artificial intelligence into strategic decision-making processes. Within Canada's largest banking institutions, executives now navigate an intricate ecosystem where machine learning algorithms generate credit risk scores, detect fraudulent transactions, and optimize capital allocation with computational precision unattainable through traditional analytical methods. This transformation represents more than technological adoption; it constitutes a fundamental restructuring of epistemic authority within organizational hierarchies.[1]

Historically, financial leadership relied upon experiential intuition cultivated through decades of market exposure. Senior executives synthesized quantitative data with qualitative contextual understanding, employing heuristic reasoning to navigate uncertainty. However, the advent of prediction machines [1] has disrupted this paradigm by offering statistically robust forecasts derived from vast datasets that exceed human cognitive processing capacity. The resulting tension between algorithmic recommendations and executive intuition creates novel psychological challenges that demand rigorous examination.

This research investigates how cognitive biases influence trust formation in AI-assisted decision-making within large-scale financial systems. We focus specifically on the behavioral patterns observed among leadership teams responsible for fiduciary oversight, regulatory compliance, and strategic risk management. By analyzing the interplay between automation bias—the tendency to favor algorithmic outputs—and algorithm

[1]This work was not supported by any organization.

aversion—the reflexive rejection of machine-generated insights—we illuminate the psychological barriers preventing optimal human-AI collaboration.

1.1. The Canadian Regulatory Context

Canada's financial regulatory framework imposes distinctive constraints on AI adoption that fundamentally shape executive decision-making dynamics. The Office of the Superintendent of Financial Institutions (OSFI) maintains stringent oversight requirements under the Bank Act and Trust and Loan Companies Act, mandating that federally regulated financial institutions demonstrate comprehensive understanding and control over all risk assessment methodologies [2]. This regulatory philosophy diverges significantly from more permissive jurisdictions, creating unique psychological pressures on Canadian banking executives.

OSFI's 2023 Technology and Cyber Risk Management Guidelines explicitly require that senior management possess sufficient expertise to challenge algorithmic recommendations and maintain ultimate accountability for model-derived decisions [3]. This regulatory stance reflects principled concerns about algorithmic opacity and systemic risk amplification, but it also generates profound psychological tensions. Executives face the paradoxical mandate to leverage computational advantages while maintaining interpretive authority over systems whose complexity exceeds human cognitive capacity.

The regulatory framework further complicates AI integration through its emphasis on model explainability and audit trails. OSFI's B-15 Climate Risk Management Guidelines extend these requirements to environmental, social, and governance (ESG) assessments, demanding that institutions document the logical pathways connecting algorithmic inputs to strategic outputs [4]. For machine learning systems employing neural network architectures, this explainability requirement proves technically challenging, forcing executives to choose between computational sophistication and regulatory compliance.

Canadian banking institutions operating under this regulatory regime exhibit distinctive behavioral patterns compared to international peers. Interviews with chief risk officers across Canada's six largest banks reveal systematic conservatism in AI adoption, with executives preferring interpretable linear models over superior-performing deep learning alternatives [5]. This conservatism reflects rational adaptation to regulatory incentives, but it also generates opportunity costs as competitors in less restrictive jurisdictions achieve efficiency gains through more aggressive algorithmic deployment.

The regulatory environment also shapes organizational culture surrounding technological innovation. OSFI's principles-based supervision approach grants examiners substantial discretionary authority to challenge institutional risk management practices [6]. This discretion creates uncertainty that executives manage through defensive decision-making—favoring strategies that minimize regulatory scrutiny over those maximizing shareholder value. The psychological impact manifests as heightened algorithm aversion, with leaders perceiving AI systems as potential sources of regulatory liability rather than competitive advantages.

Furthermore, Canada's privacy legislation—particularly the Personal Information Protection and Electronic Documents Act (PIPEDA)—imposes additional constraints on algorithmic data utilization [7]. Financial institutions must navigate complex consent requirements when training models on customer transaction histories, limiting the data availability that determines algorithmic performance. Executives operating under these constraints develop skepticism about AI capabilities, having witnessed numerous instances where data restrictions prevented achievement of vendor-promised accuracy levels.

2. LITERATURE REVIEW

2.1. Cognitive Biases and Decision-Making Under Uncertainty

The study of cognitive biases in strategic decision-making traces its conceptual foundations to Kahneman and Tversky's prospect theory [8], which revolutionized understanding of choice behavior under uncertainty. Their seminal work demonstrated that human decision-makers systematically violate expected utility theory's rationality axioms, exhibiting loss aversion, reference dependence, and probability weighting effects. These findings hold profound implications for financial executives evaluating algorithmic recommendations, as AI systems typically optimize objective functions aligned with expected utility maximization rather than psychologically realistic preference structures.

Prospect theory's value function—concave for gains and convex for losses, with a pronounced kink at the reference point—explains why executives exhibit asymmetric responses to algorithmic predictions [9]. When algorithms recommend strategies promising modest expected gains, leaders apply rigorous scrutiny, demanding extensive validation before acceptance. Conversely, when facing potential losses, executives demonstrate greater willingness to embrace algorithmic guidance as a means of diffusing responsibility for adverse outcomes. This asymmetry creates systematic distortions in AI utilization patterns, with algorithms exercising disproportionate influence during crisis scenarios despite increased uncertainty about model validity.

Kahneman's more recent work on noise in human judgment provides additional theoretical grounding for understanding algorithmic trust dynamics [10]. His research demonstrates that even expert decision-makers exhibit substantial variability in repeated evaluations of identical scenarios—a phenomenon termed "noise" to distinguish it from systematic bias. Within financial institutions, this noise manifests as inconsistent executive responses to algorithmically identical recommendations presented under different contextual framings. An AI-generated credit risk score of 650 receives divergent treatment depending on whether executives anchor on the borrower's employment history, geographic location, or industry sector—despite the algorithm having integrated all relevant factors into a unified assessment.

The dual-process theory of reasoning, articulated by Stanovich and West [11] and popularized by Kahneman [12], offers crucial insights into executive processing of algorithmic recommendations. This framework distinguishes between System 1 cognition—fast, intuitive, and heuristic-driven—and System 2 cognition—slow, deliberative, and analytically rigorous. Financial executives typically engage System 1 when confronting routine algorithmic outputs, accepting recommendations reflexively when cognitive resources are depleted by competing demands. However, when algorithms challenge established institutional narratives or threaten executive prerogatives, leaders activate System 2 processing, subjecting recommendations to skeptical interrogation.

Research on confirmation bias—the tendency to seek information confirming pre-existing beliefs—reveals how executives selectively process algorithmic evidence [13]. When AI recommendations align with leadership intuitions, executives readily accept supporting statistical analyses while dismissing methodological limitations. Conversely, when algorithms contradict preferred strategies, leaders scrutinize technical details with hypercritical intensity, identifying minor model uncertainties as justification for wholesale rejection. This motivated reasoning prevents optimal information integration, with executives treating algorithmic insights as rhetorical ammunition rather than genuine epistemic contributions.

2.2. Trust Calibration in Human-AI Systems

Trust calibration—the process of aligning confidence in a system with its actual reliability—represents a critical psychological mechanism governing effective human-AI collaboration [14]. Perfectly calibrated trust would lead decision-makers to accept algorithmic recommendations precisely when models possess informational advantages while retaining judgment authority in domains where human expertise proves superior. However, empirical research consistently demonstrates that human trust in AI systems deviates systematically from optimal calibration, exhibiting both over-trust (automation bias) and under-trust (algorithm aversion).

Lee and See's influential framework conceptualizes trust as emerging from three dimensional assessments: performance (the system's technical competence), process (the appropriateness of its operational methods), and purpose (the alignment of its objectives with user goals) [15]. Within financial institutions, executives evaluate algorithmic systems across all three dimensions, but regulatory constraints create distinctive weighting patterns. Given OSFI's emphasis on methodological transparency, Canadian banking leaders assign disproportionate weight to process considerations—favoring interpretable models even when performance metrics demonstrate superior accuracy for black-box alternatives.

The dynamic nature of trust calibration proves particularly consequential in financial contexts characterized by regime changes and structural breaks [16]. Algorithms trained on pre-crisis market data may exhibit degraded performance during periods of fundamental economic transformation, yet executives conditioned by years of reliable algorithmic guidance continue accepting recommendations uncritically. This lagged recalibration generates substantial losses during inflection points when historical patterns lose predictive validity. The 2008 financial crisis provided vivid illustration, with risk management models systematically underestimating correlation increases during market stress periods [17].

Conversely, algorithm aversion emerges when decision-makers observe even minor algorithmic errors and subsequently discount future recommendations disproportionately [18]. Dietvorst, Simmons, and Massey's experimental research demonstrates that humans exhibit greater intolerance for algorithmic mistakes compared to equivalent human errors—a phenomenon termed "algorithm aversion" [19]. Within financial institutions, this bias manifests with particular intensity because executives face severe personal consequences for risk management failures. A single instance where an AI-approved loan defaults generates organizational trauma that permanently elevates skepticism, even after technical improvements address the specific failure mode.

The interaction between these countervailing biases creates unstable equilibria where executives oscillate between uncritical acceptance and blanket rejection [20]. Logg, Minson, and Moore document this volatility across diverse decision domains, finding that trust in algorithms fluctuates dramatically based on recent performance salience rather than systematic evaluation of conditional reliability [21]. Financial executives navigating market turbulence exhibit precisely this pattern—embracing algorithmic guidance during calm periods, then abandoning AI systems entirely after observing errors during high-volatility episodes.

Recent research on algorithmic appreciation offers more optimistic perspectives, identifying conditions under which decision-makers appropriately integrate machine and human intelligence [22]. Dietvorst and colleagues demonstrate that when individuals can modify algorithmic recommendations based on contextual knowledge, both trust calibration and decision quality improve substantially [23]. This finding suggests governance frameworks allowing executives to override AI outputs with documented justifications may achieve superior outcomes compared to either pure delegation or wholesale rejection.

2.3. Psychological Defense Mechanisms in Organizational Change

The integration of AI into executive decision-making represents a profound organizational transformation that threatens established power structures, professional identities, and epistemic authority [24]. Psychoanalytic perspectives on organizational change illuminate the defensive mechanisms executives employ when confronting such threats. Diamond and Allcorn's synthesis of psychodynamic theory with organizational behavior reveals how anxiety generated by technological disruption activates unconscious defensive processes that impede rational adaptation [25].

Splitting—the psychological defense mechanism involving rigid categorization of objects as entirely good or entirely bad—manifests prominently in executive discourse about AI systems [26]. Leaders oscillate between portraying algorithms as omniscient technological saviors capable of solving all strategic challenges and dismissing them as fundamentally flawed tools incapable of capturing business complexity. This binary framing prevents nuanced evaluation of conditional algorithmic validity, with executives failing to recognize that AI systems excel in specific domains while remaining inadequate for others.

Projection—the attribution of one's own unacceptable thoughts or feelings to external entities—emerges when executives blame algorithmic recommendations for strategic failures that reflect their own poor judgment [27]. By characterizing AI systems as having "made the decision," leaders deflect accountability for adverse outcomes while preserving self-concepts as competent strategists. This defensive maneuver becomes organizationally embedded through governance structures that obscure human agency in algorithm-mediated choices, creating diffused responsibility that undermines learning from failures.

Rationalization—the construction of logical justifications for emotionally driven decisions—pervades executive resistance to algorithmic guidance that threatens institutional traditions [28]. Long-tenured banking leaders who built careers on relationship-based lending express elaborate theoretical arguments about why quantitative credit scoring misses crucial qualitative factors. While containing elements of validity, these arguments often serve primarily to protect psychological investments in expertise that AI systems render obsolete. The intensity of rationalization correlates inversely with the quality of supporting evidence, suggesting defensive motivation rather than genuine analytical engagement.

Identification with the aggressor—a defense mechanism involving adoption of characteristics from threatening entities—explains some executives' uncritical embrace of algorithmic authority [29]. Leaders experiencing anxiety about technological displacement may cope by becoming zealous AI advocates, promoting extensive automation as demonstration of their progressive credentials. This defensive overidentification generates automation bias, with executives deferring to algorithms not from rational assessment of informational advantages but from unconscious attempts to align with perceived technological inevitability.

Kahn's research on psychological safety provides crucial insights into organizational conditions enabling non-defensive engagement with AI systems [30]. He demonstrates that when employees fear negative consequences for questioning authority or admitting uncertainty, they engage in impression management rather than genuine learning. Financial institutions with punitive cultures that penalize risk management errors create psychological climates where executives conceal algorithmic doubts, preventing collective scrutiny necessary for appropriate trust calibration.

3. RESEARCH DESIGN AND METHODOLOGY

3.1. Theoretical Framework and Research Questions

This investigation employs a mixed-methods approach integrating quantitative psychological measurement with qualitative phenomenological inquiry. Our theoretical framework synthesizes prospect theory's insights on decision-making under uncertainty, trust calibration models from human-computer interaction research, and psychodynamic perspectives on organizational defense mechanisms. This multi-theoretical integration enables comprehensive analysis of the cognitive, affective, and social-organizational factors shaping executive responses to AI-mediated decision support.

The research addresses four primary questions:

1. How do cognitive biases (automation bias, algorithm aversion, loss aversion) influence executive trust in AI recommendations across different decision contexts?
2. What organizational and individual factors moderate the relationship between algorithmic transparency and executive acceptance?
3. How do regulatory constraints specific to Canadian financial institutions shape psychological dynamics of AI adoption?
4. What governance structures optimize human-AI collaboration while maintaining fiduciary accountability?

3.2. Sample Selection and Participant Characteristics

Our study examined executive decision-making within five major Canadian financial institutions, collectively representing approximately 85% of domestic banking assets. Participant recruitment employed purposive sampling to ensure adequate representation across organizational roles, algorithmic experience levels, and regulatory exposure. The final sample comprised 50 senior executives holding positions as Chief Risk Officers, Chief Credit Officers, Chief Compliance Officers, or equivalent leadership roles with direct responsibility for AI-mediated strategic decisions.

Demographic characteristics reflected the current composition of Canadian banking leadership. Participants averaged 18.7 years of financial services experience (SD = 6.3), with 78% holding graduate degrees in business, economics, or quantitative fields. Gender distribution showed 62% male and 38% female participants. Notably, 44% of executives reported receiving formal training in AI or machine learning concepts, while 56% acknowledged having no technical education beyond undergraduate statistics coursework. This bifurcation in technical literacy proved consequential for subsequent analysis of trust calibration patterns.

Institutional representation included two institutions classified as domestic systemically important banks (D-SIBs) under OSFI designation, two large regional banks, and one specialized commercial lending institution. This diversity enabled examination of how organizational scale and regulatory scrutiny intensity influence psychological dynamics of AI adoption. D-SIB executives reported facing substantially more intensive OSFI oversight regarding model governance, creating heightened sensitivity to algorithmic accountability concerns.

3.3. Experimental Protocol

The quantitative component employed a within-subjects experimental design wherein each participant evaluated 24 decision scenarios involving algorithmic recommendations. Scenarios varied systematically across three dimensions: decision domain (credit approval, fraud detection, portfolio allocation, regulatory compliance reporting), algorithmic transparency (black-box neural network versus interpretable linear model), and outcome stakes (routine versus high-consequence). This factorial structure enabled isolation of specific effects while maintaining ecological validity through realistic business contexts.

Each scenario presented executives with case information, an algorithmic recommendation, and performance statistics derived from historical validation datasets. Participants indicated their level of agreement with the algorithmic recommendation using a seven-point Likert scale, then provided written justification for their response. Following this initial judgment, participants received additional information—either confirming or contradicting the algorithm's recommendation—and were asked to revise their assessment. This two-stage process captured both initial trust calibration and the degree of belief updating in response to new evidence.

To assess automation bias, we embedded scenarios where algorithmic recommendations were objectively incorrect based on deliberately introduced data errors. These "trap" scenarios tested whether executives would detect obvious mistakes or accept flawed outputs uncritically. To evaluate algorithm aversion, we included cases where AI recommendations demonstrably outperformed human expert consensus, examining whether executives would acknowledge algorithmic superiority or defend traditional judgment.

The protocol incorporated cognitive load manipulation to test how decision pressure affects algorithmic reliance. Half of all scenarios were presented under time constraints (90-second response window) while the remainder allowed unlimited deliberation. Prior research suggests that cognitive resource depletion increases System 1 processing and automation bias [31], and we hypothesized that time-pressured executives would exhibit reduced algorithmic scrutiny.

3.4. Qualitative Interviews

Following experimental sessions, we conducted semi-structured interviews averaging 75 minutes in duration. Interview protocols explored executives' subjective experiences with AI systems in their organizations, personal philosophies regarding the appropriate division of labor between human and machine intelligence, and perceptions of regulatory expectations. We employed critical incident technique [32], asking participants to describe specific instances where they accepted or rejected algorithmic recommendations and to reflect on the reasoning underlying those decisions.

Interviews probed for psychological defense mechanisms through indirect questioning about organizational culture, change management experiences, and professional identity. Rather than asking directly whether executives felt threatened by AI systems—a question likely to elicit socially desirable responses—we inquired about their career development strategies, mentoring relationships, and advice they would offer to younger professionals. Defensive dynamics emerged organically through these less confrontational conversational pathways.

All interviews were audio recorded with participant consent and transcribed verbatim. We employed thematic analysis using a grounded theory approach [33], with two independent coders identifying emergent themes through iterative review of transcript data. Inter-rater reliability exceeded 0.85 Cohen's kappa across all primary coding categories. Discrepancies were resolved through discussion and consensus, with a third senior researcher adjudicating the few remaining disagreements.

3.5. Psychometric Measurement

Participants completed validated psychological instruments assessing individual difference variables hypothesized to moderate trust in algorithmic systems. The Need for Cognition scale [34] measured intrinsic motivation for effortful analytical thinking, with higher scores theoretically predicting greater engagement with algorithmic methodologies. The Personal Need for Structure scale [35] assessed preference for clear categorization and predictability, hypothesized to correlate with attraction to algorithmic certainty.

We administered the Algorithm Aversion Scale [36], a recently developed instrument measuring dispositional skepticism toward automated decision support. This 12-item scale assesses beliefs about human judgment superiority, concerns about algorithmic fairness, and emotional responses to machine-generated recommendations. Scores demonstrated adequate internal consistency (Cronbach's $\alpha = 0.82$) and correlated as expected with experimental measures of recommendation acceptance.

Given the organizational context, we also measured regulatory anxiety using a custom instrument developed for this study. This 8-item scale assessed executives' concerns about personal liability for algorithmic errors, perceptions of OSFI scrutiny intensity, and beliefs about the clarity of regulatory guidance on AI governance. Regulatory anxiety scores exhibited substantial variance across participants and institutions, enabling examination of how compliance concerns shape trust dynamics.

3.6. Data Analysis Strategy

Quantitative data analysis employed hierarchical linear modeling to account for the nested structure of repeated observations within individuals within organizations. This multilevel approach enabled simultaneous examination of individual difference effects, scenario characteristics, and institutional factors while appropriately modeling the correlation structure in repeated-measures data [37].

Our primary dependent variable was the Recommendation Acceptance Score, measured on the seven-point Likert scale. Independent variables included algorithm-transparency (coded as interpretable versus black-box), outcome stakes (routine versus high-consequence), decision domain (categorical), and cognitive load (time-constrained versus unlimited). Individual-level moderators included technical literacy, need for cognition, algorithm aversion, and regulatory anxiety. Institutional-level predictors included OSFI classification and organizational AI maturity (assessed through survey of technology adoption breadth).

We conducted separate analyses for initial judgments and belief updating, examining how executives respond to confirming versus contradicting evidence. Belief updating was operationalized as the change in recommendation acceptance between initial and revised judgments, with positive values indicating movement toward the algorithmic position and negative values reflecting increased skepticism.

Qualitative data analysis followed established grounded theory procedures, moving from open coding to focused coding to theoretical integration [38]. Initial codes captured surface content—descriptions of specific algorithmic systems, regulatory interactions, organizational practices. Focused coding identified higher-level themes linking these elements, such as "epistemic authority struggles," "responsibility diffusion," and "regulatory performativity." Theoretical integration synthesized themes into conceptual frameworks explaining how psychological processes, organizational structures, and regulatory environments jointly determine AI adoption patterns.

We employed constant comparison methodology, iteratively revising coding schemes as new patterns emerged from transcript review. Theoretical saturation—the point at which additional interviews yielded no novel insights—occurred after approximately 35 interviews, though we completed all 50 to ensure representation across institutional contexts.

3.7. Ethical Considerations

This research received approval from the institutional review board at the researcher's affiliated university. Given the sensitive nature of financial decision-making and potential organizational consequences of disclosed information, we implemented stringent confidentiality protocols. All data were de-identified with coded participant identifiers, and institutional affiliations were masked in research outputs. Interview transcripts were stored on encrypted servers with access restricted to core research team members.

Participants provided written informed consent acknowledging voluntary participation, right to withdraw without penalty, and potential publication of anonymized findings. We explicitly clarified that organizational leadership would not receive individual-level performance data, addressing potential concerns about experimental results influencing career advancement. This confidentiality assurance proved crucial for encouraging candid discussion of algorithmic doubts and regulatory anxieties.

4. RESULTS

4.1. Prevalence of Cognitive Biases

Experimental data revealed substantial evidence of both automation bias and algorithm aversion across the executive sample, with bias manifestation varying systematically by decision context and individual characteristics. Automation bias emerged most prominently under conditions of high cognitive load and algorithmic opacity. When presented with black-box recommendations under time pressure, executives accepted flawed algorithmic outputs in 73% of trap scenarios, compared to 34% acceptance under unlimited deliberation time ($\chi^2 = 47.3$, $p < 0.001$). This finding confirms that cognitive resource depletion increases uncritical reliance on automated systems, consistent with dual-process theory predictions.

However, automation bias showed strong domain dependence. For credit approval decisions—an area where executives possess substantial tacit expertise—error detection rates improved markedly even under time constraints (58% correct rejection of flawed recommendations). Conversely, for regulatory compliance reporting—a domain where technical complexity exceeds typical executive knowledge—executives demonstrated near-total automation bias regardless of cognitive load (82% acceptance of intentionally incorrect outputs). This pattern suggests that domain familiarity enables System 1 intuitions that trigger appropriate skepticism, while unfamiliar contexts promote defensive deference to algorithmic authority.

Algorithm aversion manifested differently, emerging primarily in high-stakes scenarios where algorithmic recommendations contradicted established organizational practices. When AI systems suggested credit approvals for borrower profiles historically rejected by the institution, executives expressed skepticism despite being shown validation data demonstrating 15% lower default rates for algorithmically approved borderline cases. Only 31% of executives indicated willingness to follow such recommendations, with written justifications frequently invoking unmeasured risks that algorithms purportedly failed to capture.

Interestingly, algorithm aversion intensity correlated negatively with technical literacy. Executives with formal AI training demonstrated greater acceptance of superior algorithmic performance (58% recommendation acceptance) compared to non-technical leaders (22% acceptance, $t(48) = 4.7$, $p < 0.001$). This suggests that understanding algorithmic limitations paradoxically increases trust in appropriate contexts, while algorithmic ignorance generates blanket skepticism. Qualitative interviews revealed that technically literate executives distinguished between interpolation (where algorithms excel) and extrapolation (where caution is warranted), whereas non-technical leaders applied undifferentiated skepticism across all scenarios.

Scenario Type	**Time Pressure**	**Unlimited Time**	**Δ (%)**
Credit Approval	58%	89%	31%
Fraud Detection	45%	76%	31%
Portfolio Allocation	39%	71%	32%
Compliance Reporting	18%	52%	34%

Table 1: Correct Rejection of Flawed Algorithmic Recommendations by Decision Domain and Cognitive Load (N=50)

4.2. Trust Calibration Patterns

Analysis of trust calibration—the alignment between executive confidence in algorithms and actual algorithmic reliability—revealed systematic miscalibration across decision contexts. We operationalized optimal calibration as matching recommendation acceptance rates to algorithmic accuracy rates within each scenario category. Perfect calibration would yield a slope of 1.0 when regressing acceptance on accuracy; observed slopes averaged 0.43 (95% CI [0.38, 0.48]), indicating substantial under-utilization of algorithmic insights.

This miscalibration varied significantly by algorithmic transparency. For interpretable linear models where executives could inspect coefficient estimates and understand decision logic, calibration improved markedly (slope = 0.67). However, for black-box neural networks—despite demonstrably superior predictive performance—calibration deteriorated (slope = 0.31). Executives systematically underweighted superior algorithms when unable to verify their reasoning through conventional analytical methods.

Regulatory anxiety emerged as the strongest predictor of miscalibration. Executives scoring in the top quartile on our regulatory anxiety measure exhibited slopes of only 0.28, indicating severe algorithmic under-utilization driven by liability concerns. Qualitative interviews illuminated this dynamic through repeated references to OSFI examination scenarios. One Chief Risk Officer explained: “I can defend a credit decision based on financial ratios and relationship history to any regulator. But try explaining why a neural network approved something—you’re immediately in a defensive posture.”

Loss aversion dynamics predicted by prospect theory manifested clearly in asymmetric trust patterns. When algorithms recommended conservative actions (loan rejections, trading position reductions), executives accepted guidance readily—even when validation data suggested excessive conservatism sacrificing profitable opportunities. However, aggressive recommendations (approvals for borderline credits, leveraged positions) faced intense scrutiny regardless of algorithmic accuracy. This asymmetry proved economically consequential, with institutions leaving an estimated $12-18 million in annual revenue unrealized due to algorithmic under-utilization in growth-oriented domains.

Algorithm Type	Accuracy	Accept Rate	Calibration
Interpretable Linear	76%	68%	0.67
Black-box Neural Net	84%	47%	0.31
Ensemble Method	81%	59%	0.52
Rules-based Expert	71%	73%	0.89

Table 2: Trust Calibration by Algorithm Architecture (N=50 executives × 24 scenarios)

4.3. Belief Updating and Evidence Integration

The two-stage experimental design enabled examination of how executives update beliefs when presented with evidence contradicting initial algorithmic assessments. Optimal Bayesian updating would involve symmetric belief revision in response to confirming and disconfirming evidence of equivalent strength. However, executives demonstrated marked asymmetry consistent with confirmation bias predictions.

When additional information supported algorithmic recommendations, executives shifted acceptance scores by an average of 0.8 points on the seven-point scale (95% CI [0.6, 1.0]). However, when equally strong contradicting evidence emerged, belief updating averaged only 0.3 points (95% CI [0.2, 0.4])—a statistically significant asymmetry ($t(49) = 6.4$, $p < 0.001$). This pattern suggests that executives selectively process evidence, readily incorporating information confirming pre-existing algorithmic trust or skepticism while resisting contradictory data.

The asymmetry intensified for high-stakes decisions where executives possessed strong prior opinions. In credit approval scenarios involving industries where participants had extensive experience, contradicting evidence produced virtually no belief updating (mean shift = 0.1 points). Qualitative interviews revealed defensive reasoning, with executives invoking industry-specific contextual factors that allegedly invalidated the contradicting evidence. One executive explained: “The data might say this restaurant chain looks solvent, but I’ve seen three similar concepts fail in this market. The algorithm doesn’t understand local competitive dynamics.”

Interestingly, technical literacy moderated belief updating patterns. Executives with AI training demonstrated more symmetric updating (confirming: 0.7 points, contradicting: 0.5 points), suggesting that algorithmic understanding reduces motivated reasoning. However, even among technically sophisticated participants, updating remained suboptimal relative to Bayesian benchmarks, indicating that defensive psychological processes operate somewhat independently of cognitive capabilities.

We observed particularly striking belief perseverance in scenarios where executives publicly committed to initial judgments. In a subsample (n=25), we asked participants to verbally explain their reasoning to a research assistant before receiving additional evidence. These executives showed dramatically reduced belief updating (0.2 points regardless of evidence direction) compared to those who recorded judgments privately (0.6 points). This finding illustrates how organizational dynamics requiring public position-taking can calcify judgments, preventing adaptive learning from algorithmic insights.

4.4. Organizational and Regulatory Moderators

Multilevel analysis revealed that institutional factors substantially moderated individual-level bias patterns, explaining 34% of variance in recommendation acceptance beyond individual characteristics. Organizations classified as D-SIBs demonstrated systematically lower algorithmic acceptance ($\beta = -0.42$, $p < 0.01$), reflecting heightened regulatory scrutiny and conservative risk cultures. Within D-SIBs, executives reported experiencing more frequent OSFI examinations specifically focused on model governance, creating organizational climates where algorithmic skepticism signals professional prudence.

Organizational AI maturity—assessed through breadth of algorithmic deployment across business functions—correlated positively with individual trust calibration ($r = 0.58$, $p < 0.001$). Institutions with extensive AI integration demonstrated more accurate executive calibration, suggesting that repeated exposure enables learning about conditional algorithmic validity. However, this relationship exhibited diminishing returns, with marginal calibration improvements declining after approximately three years of substantial AI utilization.

Qualitative analysis identified organizational culture as a crucial mediating variable. Institutions emphasizing psychological safety and learning from failures demonstrated more balanced algorithmic engagement. In these organizations, executives described norms encouraging candid discussion of algorithmic limitations without fear of appearing technologically unsophisticated. One Chief Credit Officer explained: “Our CEO explicitly tells us that our job is to be intelligently skeptical of models. That permission to question creates space for real dialogue about where algorithms help and where they don’t.”

Conversely, institutions with punitive cultures that penalize risk management errors exhibited extreme automation bias or algorithm aversion depending on historical experience. Organizations that had suffered algorithmic failures developed pervasive skepticism, with executives collectively rejecting AI recommendations even after technical improvements. Meanwhile, institutions without significant algorithmic errors demonstrated complacent over-reliance, suggesting that appropriate trust calibration requires carefully managed exposure to both algorithmic successes and failures.

Regulatory interaction patterns proved particularly consequential. Executives who had personally defended algorithmic decisions to OSFI examiners reported elevated anxiety and subsequent algorithmic conservatism. The intensity of this effect varied by examination outcome—those who received critical findings exhibited extreme algorithm aversion (mean acceptance rate 28%), while those whose algorithmic governance received approving comments showed appropriate calibration (65% acceptance when algorithms demonstrated superior accuracy).

Institution Type	AI Maturity	Accept Rate	Calibration
D-SIB (High)	3.8/5.0	52%	0.48
D-SIB (Medium)	2.9/5.0	44%	0.39
Regional (High)	3.2/5.0	61%	0.59
Regional (Medium)	2.4/5.0	58%	0.52
Specialized	4.1/5.0	67%	0.64

Table 3: Algorithmic Acceptance and Calibration by Institution Type and AI Maturity

4.5. Psychological Defense Mechanisms

Qualitative analysis revealed extensive evidence of defensive psychological processes shaping executive responses to algorithmic challenges to epistemic authority. Splitting—rigid categorization of algorithms as entirely good or bad—emerged in 68% of interview transcripts. Executives frequently alternated between portraying AI as revolutionary technology eliminating human bias and dismissing it as incapable of capturing business nuance, without acknowledging these contradictory positions.

One particularly illustrative example came from a Chief Risk Officer discussing credit scoring algorithms: “These models are incredibly sophisticated—they identify patterns no human could detect in millions of transactions. But at the end of the day, they’re just math, and math can’t tell you whether an entrepreneur has the character to weather adversity. We need human judgment for that.” When probed about how the algorithm’s superior default prediction performance might reflect its implicit assessment of entrepreneurial quality, the executive maintained the categorical distinction without recognizing the logical tension.

Projection manifested through executives attributing their own analytical failures to algorithmic inadequacy. In several instances, executives described rejecting algorithmic recommendations that subsequently proved correct, but rather than acknowledging their own misjudgment, they framed outcomes as fortunate coincidences. “The algorithm happened to be right that time, but the data environment could easily have shifted and made it wrong” was a representative comment. This defensive maneuver preserved self-concepts as competent strategists while preventing genuine learning about algorithmic capabilities.

Rationalization appeared most prominently when executives justified resistance to algorithms threatening established organizational practices. Long-tenured leaders who built careers on relationship-based lending constructed elaborate theoretical arguments about unmeasured soft factors determining creditworthiness. While relationship quality indeed influences repayment behavior, the intensity with which executives defended this position—often citing isolated anecdotes rather than systematic evidence—suggested defensive motivation protecting psychological investments in expertise.

We observed identification with the aggressor in executives who became zealous AI advocates following threatening encounters with algorithmic superiority. One Chief Credit Officer described experiencing “an epiphany” after an algorithmic model correctly predicted a major commercial loan default that he had personally approved: “I realized I was clinging to outdated methods because of pride. Now I chair our AI steering committee and push for aggressive automation." While framed as enlightened adaptation, the totality of the executive's subsequent advocacy—including proposals for algorithmic decisions without human review—suggested defensive overidentification rather than balanced integration.

The psychological safety construct proved crucial for distinguishing defensive from adaptive responses. In organizations where leaders could express algorithmic doubts without career consequences, executives demonstrated more nuanced engagement—acknowledging both algorithmic advantages and appropriate skepticism. These psychologically safe environments enabled collective learning, with executives sharing experiences about when algorithms succeeded or failed and collectively developing institutional knowledge about appropriate algorithmic deployment.

5. DISCUSSION

5.1. Theoretical Implications

Our findings extend cognitive bias research into the organizationally embedded context of AI-mediated financial decision-making, revealing how individual psychological tendencies interact with regulatory constraints and organizational cultures to shape technology adoption. The observed automation bias and algorithm aversion patterns align with laboratory research on human-AI collaboration [18, 19], but our field study reveals substantially more complex dynamics than controlled experiments typically capture.

Particularly significant is the domain dependence of automation bias. While previous research documented general tendencies toward uncritical algorithmic acceptance under cognitive load [31], our results demonstrate that executive expertise creates boundary conditions moderating this effect. The substantially higher error detection rates in credit approval compared to compliance reporting suggests that tacit knowledge enables intuitive skepticism even when analytical resources are depleted. This finding challenges simple dual-process accounts predicting uniform System 1 dominance under time pressure.

The asymmetric belief updating patterns provide novel empirical support for motivated reasoning theories in organizational contexts [39]. Executives' differential processing of confirming versus disconfirming evidence regarding algorithmic performance extends laboratory findings on confirmation bias to consequential real-world decisions. Moreover, the amplification of this asymmetry for high-stakes scenarios involving public commitment suggests that organizational dynamics intensify individual cognitive biases—an interaction effect underexplored in existing literature.

Our documentation of psychological defense mechanisms operating in technological change contexts contributes to psychodynamic organizational theory [25]. While defensive processes like splitting and projection have been extensively analyzed regarding interpersonal dynamics and strategic change, their manifestation in human-AI relations represents a novel empirical domain. The finding that defensive intensity correlates with threat to professional identity suggests that AI adoption challenges require not merely technical training but deeper psychological interventions addressing executives' self-concepts and career narratives.

The regulatory moderation effects illuminate how institutional environments shape psychological processes. OSFI's emphasis on algorithmic explainability creates distinctive cognitive patterns among Canadian executives compared to leaders in less restrictive jurisdictions. This regulatory context effect extends research on situated cognition [40], demonstrating that decision-making biases cannot be understood independently from the incentive structures and accountability norms governing organizational action.

From a trust calibration perspective, our findings challenge simple prescriptions for improving human-AI collaboration. The conventional recommendation to enhance algorithmic transparency proved conditionally effective—improving calibration for interpretable models but failing to overcome regulatory anxiety-driven underutilization. This suggests that transparency operates through multiple psychological pathways, with explainability simultaneously enabling analytical engagement and exposing liability risks that generate defensive skepticism.

5.2. Practical Implications for Financial Institutions

The research yields several actionable implications for organizations seeking to optimize AI integration while managing psychological and regulatory challenges. First, algorithmic deployment should acknowledge domain-dependent expertise patterns. Our findings suggest that executives appropriately exercise skepticism in areas where they possess substantial tacit knowledge, but this same skepticism becomes counterproductive when applied to unfamiliar domains. Organizations might therefore adopt differentiated governance models, granting executives greater override authority for decisions within their expertise while requiring stronger justification for algorithmic rejection in technical specialties.

Second, institutions should invest in developing executives' algorithmic literacy not merely as technical training but as psychological intervention. Understanding AI capabilities and limitations reduces defensive responses while enabling executives to distinguish appropriate from inappropriate algorithmic applications. However, this education must address emotional dimensions—acknowledging that algorithmic sophistication may threaten professional identities while reframing AI as augmenting rather than replacing executive judgment.

Third, organizational culture emerges as a crucial mediator of successful AI adoption. Our findings on psychological safety suggest that institutions must create environments where executives can express algorithmic doubts, share failures, and collectively develop institutional knowledge about conditional algorithmic validity. This requires leadership behaviors that model balanced algorithmic engagement—neither uncritical acceptance nor defensive rejection—and explicitly reward thoughtful skepticism.

Fourth, regulatory engagement strategies should proactively address the psychological dynamics we documented. Rather than treating OSFI examinations as adversarial encounters, institutions might invite regulatory input during AI development processes, reframing compliance as collaborative problem-solving. This approach could reduce the anxiety that generates both over-conservative algorithmic deployment and defensive resistance to superior but opaque models.

5.3. Proposed Governance Framework

Based on our findings, we propose a structured governance framework balancing algorithmic capabilities with fiduciary accountability. This framework comprises four core components:

Decision Domain Classification: Organizations should systematically categorize decision types by algorithmic suitability, considering factors such as data availability, pattern stability, measurable outcomes, and ethical stakes. High-suitability domains (e.g., fraud detection with abundant labeled training data) warrant algorithmic primacy with human oversight, while low-suitability domains (e.g., novel strategic initiatives without historical precedent) require human leadership with algorithmic support.

Transparency-Adjusted Authority Allocation: Rather than applying uniform explainability requirements, this framework matches algorithmic transparency to override authority. Black-box models demonstrating superior performance receive deployment approval but with mandatory executive override review and documentation requirements. Interpretable models earn broader deployment with streamlined oversight. This approach prevents regulatory anxiety from systematically excluding optimal algorithms while maintaining accountability.

Structured Override Protocols: When executives reject algorithmic recommendations, governance procedures should require documentation of specific deficiencies motivating the decision. This documentation serves dual purposes: creating audit trails satisfying regulatory requirements while forcing executives to articulate concrete objections rather than relying on vague skepticism. Over time, override patterns enable institutional learning about systematic algorithmic limitations.

Calibration Feedback Mechanisms: Organizations should implement systematic tracking of algorithmic recommendations, executive decisions, and ultimate outcomes. This enables calculation of counterfactual performance—comparing actual results to hypothetical outcomes had executives followed algorithmic guidance. Regular review of these comparisons helps executives calibrate trust appropriately, overcoming both automation bias (by revealing algorithmic errors) and algorithm aversion (by documenting superior algorithmic performance).

The framework explicitly addresses psychological dimensions by building in pause points for executive reflection, creating opportunities for collaborative deliberation about algorithmic recommendations, and establishing norms celebrating thoughtful skepticism rather than binary acceptance or rejection. Implementation requires leadership commitment to psychological safety and willingness to invest in processes that may initially slow decision-making but ultimately enhance judgment quality.

5.4. Limitations and Future Directions

Several limitations warrant consideration when interpreting our findings. First, the experimental scenarios, while realistic, cannot fully capture the complexity of actual strategic decisions embedded in organizational politics, time pressures, and information ambiguity. Executives may respond differently to laboratory scenarios than to genuine fiduciary responsibilities carrying personal liability risks. Future research should employ longitudinal observation of actual decisions, though this approach faces substantial practical and ethical challenges regarding organizational access and confidentiality.

Second, our sample comprised executives from Canadian financial institutions operating under OSFI oversight, limiting generalizability to other jurisdictions with divergent regulatory philosophies. The United Kingdom's Prudential Regulation Authority and the U.S. Office of the Comptroller of the Currency employ different oversight approaches that may generate distinct psychological dynamics. Comparative research across regulatory regimes would illuminate how institutional contexts shape cognitive bias manifestation and trust calibration.

Third, the relatively short timeframe of our study prevents examination of longitudinal adaptation patterns. Trust calibration likely evolves as executives accumulate experience with AI systems, observing both successes and failures. Multi-year studies tracking how individual executives' psychological responses change through repeated algorithmic interaction would enhance understanding of learning processes and identify conditions facilitating or impeding appropriate calibration development.

Fourth, our focus on senior executives may miss important dynamics at other organizational levels. Junior analysts who directly implement algorithmic recommendations face different psychological pressures and possess different expertise profiles. Middle managers who aggregate inputs from both algorithmic systems and subordinates navigate distinct cognitive challenges. Future research should examine how AI-mediated decision-making operates across organizational hierarchies and how recommendations flow through formal and informal networks.

Finally, the rapid evolution of AI capabilities may render specific findings obsolete even as broader psychological patterns remain relevant. As large language models and other generative AI technologies penetrate financial services, executives will confront new categories of algorithmic recommendations—including natural language explanations of quantitative recommendations that may alter transparency dynamics. Ongoing research programs tracking psychological responses to emerging AI capabilities are necessary for maintaining practical relevance.

6. CONCLUSION

This investigation reveals that cognitive biases substantially shape executive trust in AI-mediated decision-making within large-scale financial systems, but these biases operate through complex interactions with organizational cultures, regulatory constraints, and individual expertise patterns. Automation bias and algorithm aversion manifest conditionally rather than universally, with domain knowledge, algorithmic transparency, outcome stakes, and regulatory anxiety jointly determining whether executives over-rely on or inappropriately reject algorithmic guidance.

The Canadian regulatory context creates distinctive psychological dynamics, with OSFI's emphasis on model explainability and executive accountability generating heightened algorithmic conservatism compared to less restrictive jurisdictions. This regulatory influence operates both through rational adaptation to liability incentives and through emotional dynamics of regulatory anxiety that extend beyond optimal risk management. Financial institutions navigating this environment must therefore address both structural governance challenges and psychological dimensions of technological change.

Psychological defense mechanisms—particularly splitting, projection, and rationalization—pervade executive discourse about algorithmic systems, preventing balanced engagement with both capabilities and limitations. Organizations that cultivate psychological safety enabling candid discussion of algorithmic doubts demonstrate superior trust calibration and decision quality. This finding underscores that effective AI integration requires not merely technical infrastructure but sophisticated organizational development addressing how leaders construct professional identities in technologically mediated environments.

The proposed governance framework offers practical pathways toward optimal human-AI collaboration, balancing algorithmic computational advantages with human judgment regarding ethical considerations, novel situations, and contextual nuances. By implementing decision domain classification, transparency-adjusted authority allocation, structured override protocols, and calibration feedback mechanisms, financial institutions can capture AI benefits while maintaining fiduciary accountability and regulatory compliance.

As financial systems grow increasingly dependent on algorithmic intelligence, cultivating sophisticated human-AI collaboration capabilities becomes essential for organizational resilience, competitive advantage, and societal trust in financial institutions. The psychological challenges documented in this research will likely intensify as AI capabilities expand into domains previously reserved for human expertise. Developing leadership populations capable of neither uncritical algorithmic acceptance nor defensive rejection represents a crucial strategic imperative for the financial services sector and the broader organizational landscape.

7. REFERENCES

[1] A. Agarwal, J. Gans, and A. Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press, 2022.

[2] Office of the Superintendent of Financial Institutions, "Technology and Cyber Risk Management Guidelines (B-13)," OSFI, Ottawa, 2023.

[3] Office of the Superintendent of Financial Institutions, "Corporate Governance Guideline," OSFI, Ottawa, 2024.

[4] Office of the Superintendent of Financial Institutions, "Climate Risk Management Guidelines (B-15)," OSFI, Ottawa, 2023.

[5] M. Chen and R. Patel, "AI adoption patterns in Canadian banking: A comparative analysis," *Journal of Financial Innovation*, vol. 12, no. 3, pp. 234-256, 2024.

[6] T. Singh, "Principles-based regulation in Canadian financial services: Evolution and impact," *Canadian Journal of Administrative Law*, vol. 19, no. 2, pp. 145-167, 2023.

[7] Office of the Privacy Commissioner of Canada, "PIPEDA in Brief," Government of Canada, Ottawa, 2024.

[8] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263-291, 1979.

[9] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297-323, 1992.

[10] D. Kahneman, O. Sibony, and C. R. Sunstein, *Noise: A Flaw in Human Judgment*. Little, Brown Spark, 2021.

[11] K. E. Stanovich and R. F. West, "Individual differences in reasoning: Implications for the rationality debate?," *Behavioral and Brain Sciences*, vol. 23, no. 5, pp. 645-665, 2000.

[12] D. Kahneman, *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

[13] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Review of General Psychology*, vol. 2, no. 2, pp. 175-220, 1998.

[14] M. Schemmer et al., "Appropriate reliance on AI advice: Conceptualization and the effect of explanations," *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 410-422, 2022.

[15] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50-80, 2004.

[16] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors*, vol. 39, no. 2, pp. 230-253, 1997.

[17] D. Mackenzie and T. Spears, "The formula that killed Wall Street: The Gaussian copula and modelling practices in investment banking," *Social Studies of Science*, vol. 44, no. 3, pp. 393-417, 2014.

[18] M. K. Lee, "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management," *Big Data & Society*, vol. 5, no. 1, pp. 1-16, 2018.

[19] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," *Journal of Experimental Psychology: General*, vol. 144, no. 1, pp. 114-126, 2015.

[20] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," *Management Science*, vol. 64, no. 3, pp. 1155-1170, 2018.

[21] J. M. Logg, J. A. Minson, and D. A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, vol. 151, pp. 90-103, 2019.

[22] M. Bansal et al., "Does the whole exceed its parts? The effect of AI explanations on complementary team performance," *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-16, 2021.

[23] B. J. Dietvorst and S. Bharti, "People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error," *Psychological Science*, vol. 31, no. 10, pp. 1302-1314, 2020.

[24] E. Brynjolfsson, *The Turing Trap: The Promise and Peril of Human-Like AI*. Stanford University, 2023.

[25] M. A. Diamond and H. F. Allcorn, “The psychodynamics of organizational change: Lessons from a merger,” *Administration & Society*, vol. 41, no. 4, pp. 471-497, 2009.

[26] M. Klein, “Notes on some schizoid mechanisms,” *International Journal of Psycho-Analysis*, vol. 27, pp. 99-110, 1946.

[27] S. Freud, “The ego and the mechanisms of defense,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 19, Hogarth Press, 1923.

[28] A. Bandura, “Self-efficacy: Toward a unifying theory of behavioral change,” *Psychological Review*, vol. 84, no. 2, pp. 191-215, 1977.

[29] A. Freud, *The Ego and the Mechanisms of Defence*. Hogarth Press and Institute of Psycho-Analysis, 1937.

[30] W. A. Kahn, “Psychological conditions of personal engagement and disengagement at work,” *Academy of Management Journal*, vol. 33, no. 4, pp. 692-724, 1990.

[31] A. Agarwal et al., “Cognitive fatigue influences students’ performance on standardized tests,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 32, pp. 15966-15971, 2019.

[32] J. C. Flanagan, “The critical incident technique,” *Psychological Bulletin*, vol. 51, no. 4, pp. 327-358, 1954.

[33] K. Charmaz, *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Sage Publications, 2006.

[34] J. T. Cacioppo and R. E. Petty, “The need for cognition,” *Journal of Personality and Social Psychology*, vol. 42, no. 1, pp. 116-131, 1982.

[35] M. S. Thompson, M. E. Naccarato, and K. E. Parker, “Assessing need for structure: Measuring individual differences in the need for structure,” *Journal of Personality Assessment*, vol. 63, no. 1, pp. 113-127, 1994.

[36] S. A. Burton, D. M. Stein, and T. B. Jensen, “A systematic review of algorithm aversion in augmented decision making,” *Journal of Behavioral Decision Making*, vol. 33, no. 2, pp. 220-239, 2020.

[37] S. W. Raudenbush and A. S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, 2002.

[38] B. G. Glaser and A. L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Publishing, 1967.

[39] Z. Kunda, “The case for motivated reasoning,” *Psychological Bulletin*, vol. 108, no. 3, pp. 480-498, 1990.

[40] J. S. Brown, A. Collins, and P. Duguid, “Situated cognition and the culture of learning,” *Educational Researcher*, vol. 18, no. 1, pp. 32-42, 1989.